NACLO and you!

The sounds of spoken language

Micha Elsner,Department of Linguistics

THE OHIO STATE UNIVERSITY

What is NACLO (North American Computational Linguistics Olympiad)?

NACLO is a nationwide contest for high schoolers.

- Thought puzzles which highlight different languages from around the world and throughout history.
- No prior knowledge of languages (or computers) is required! But you might pick up some along the way.
- The NACLO competition this year will be January 25th at OSU.
- We'll be helping you prep for the competition and then proctoring the actual test when you take it.

Why do NACLO?

- Learn more about the diversity of languages world-wide
- Practice your general reasoning and problem-solving skills
- Learn more about the field of Linguistics! We work on:
  - Language Education
  - Neurology and speech disorders
  - Archaeology
  - Artificial Intelligence
  - Language and social identity
  - And more!

Sound system problems

Some problems ask you to look for patterns in how words are pronounced.

These can be very obvious, or hidden as sub-parts in problems about scripts or word endings. There are usually a couple of sound system problems every year.

Linguists call the study of sound systems **phonetics and phonology.**

Phonetics refers to how the sounds are produced and perceived…

Phonology is the way they are arranged to make up words.

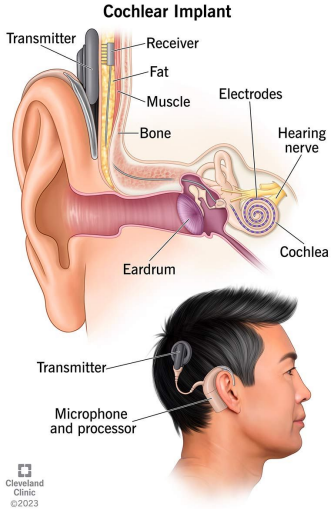
Why phonetics and phonology?

Outside NACLO, these areas are important in:

Speech pathology (img: Northwestern Univ.)

Speech recognition

(img: Apple)



Assistive technologies

(img: Cleveland Clinic)

Lesson plan

“Guided tour” of one problem in this area

- Look at the problem
- How linguistics can help
- Solution

Work on two other problems in small groups (about 20 minutes / problem)

Go over solutions

https://naclo.org/resources/problems/2021/N2021-B.pdf

Example problem

Recently, a group of researchers decided that the alphabet had too many letters… Luckily, the researchers noticed that many letters were unnecessary. For example, p and b sound a lot alike, so they decided to replace every p with a b (while every b remained unchanged). For example, *plug* would be rewritten *blug*, but *bug* would stay written as *bug*. They continued this process until they narrowed down the alphabet to just 9 letters. This writing system was named "*Duw Ulbubud*," and the media went wild. Here is one of the headlines that ran that day:

WURDZ WULL BU ZBULLUD DUVVURUDDLU

BUGUUZU UV "DUW ULBUBUD"

WURDZ WULL BU ZBULLUD DUVVURUDDLU BUGUUZU UV “DUW ULBUBUD”

Vur buddur ur vur wurzu, du drudzvurbud ulbubud wull bugu wrududg luug lugu u buzzludg zubblu uv ruddub luddurz. Wu zuzd uvu udu buzur gwuuzduud: uz ud zuzduvuud? Ugzbugd ludz uv gubbluuddz vrub buublu wu gruw ub wud du uld zuzdub.

Rewrite each of the following words in Duw Ulbubud:

words, the, fifth, squeaky, jazz, bagpipes, vertex, calming

How does “Duw Ulbubud” work?

We already know that it replaces English letters with other letters (*plug* > *blug*).

(As in any problem-solving situation, you should start writing down these patterns ASAP! NACLO problems usually reward careful notetaking.)

English	**Duw Ulbubud**
p, b	b
l	l
g	g
u	u

Guessing some more words

WURDZ WULL BU ZBULLUD DUVVURUDDLU BUGUUZU UV “DUW ULBUBUD”

words will be spelled differently because of

English	Duw Ulbubud
p, b	b
l	l
g	g
a, e, i, o, u	u

English	Duw Ulbubud
n, t, d	d
w	w
s, z	z
r	r
f, v	v

My first readthrough

Vur buddur ur vur wurzu, du drudzvurbud ulbubud wull bugu

for better or for worse, xx xxxxxxxxxx alphabet will xxxx

wrududg luug lugu u buzzludg zubbl**u** uv ruddu**b** luddurz. Wu

*writing look like a puzzling suppl**y** of rando**m** letters. We*

zuzd uvu udu buzur gwuuzduud: uz ud zuzduvuud? Ugzbugd ludz

xxxx xxx xxx xxxxx xxxxxxxxx: is it xxxxxxxxx? Expect lots

uv gubbluuddz vrub buublu wu gruw ub wud du uld zuzdub.

of complaints from people who grew up with the old system.

How can linguistics help?

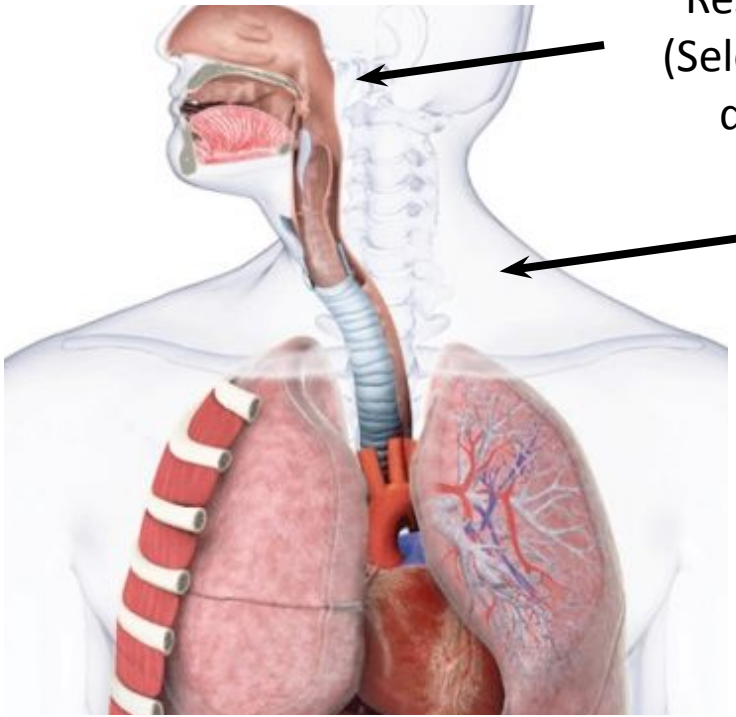
You can solve the Duw Ulbubud problem (and any NACLO problem) without knowing any linguistic theory at all.

But the Duw Ulbubud chart isn't *ruddub* at all. Knowing a little bit about how scientists think about sound systems can really help understand what's going on.

We'll go over three major distinctions about how sounds are made in the mouth:

- Airflow blocked, slightly open (hissing or buzzing), through the nose, totally open
- Lips, behind the teeth, or back of the mouth
- Voiced vs. voiceless

The vocal tract



Resonators (Select sound quality)

Glottis (Vibrates)

Lungs (Air source)

“Manner of articulation”: how open is the mouth?

- **Vowels:** all the way open (you can shout!)
  a, e, i, o, u
- **Hissing, buzzing and humming sounds:** Slightly open (you can sustain these)
  f, v, th, h; s, z, sh, zh; l, r
- **Nasal sounds:** through the nose (obviously)
  m, n, ng
- **Stops:** full closure (you can’t sustain these)
  p, b, t, d, k, g
- **(Secret 5th option: glides):** share features with multiple groups
  w, y

“Place of articulation”: where in the mouth?

- **At or just behind the lips (labial):**
  p, b, f, th, m, w
- **At or just behind the teeth (dental, alveolar):**
  t, d, s, z, r, n, l
- **Back of the mouth:**
  sh, zh, k, g, ng, h, y

Voiced or voiceless?

- “Voiced” sounds produced with the vocal cords vibrating.
  b, d, g, z
- “Voiceless” sounds without.
  p, t, k, s

Quick advertisement: the IPA chart

Manner of Articulation \ Place of Articulation	Bilabial		Labio-dental		Inter-dental		Alveolar		Palatal		Velar		Glottal	
Stop	p	b					t	d			k	g	ʔ	
Fricative			f	v	θ	ð	s	z	ʃ	ʒ			h	
Affricate									tʃ	dʒ				
Flap								ɾ						
Nasal		m						n				ŋ		
Lateral Liquid								l						
Retroflex Liquid								ɹ						
Glide	w̥	w								j				

State of the Glottis: Voiceless | Voiced

Back to Duw Ulbubud

What other letters do we expect “g” to replace?

English	Duw Ulbubud
m, p, b	b
n, t, d	d
g	g
a, e, i, o, u, y	u

English	Duw Ulbubud
f, v	v
s, z	z
l	l
r	r
w	w

My second readthrough

Vur buddur ur vur wurzu, du drudzvurbud ulbubud wull bugu

for better or for worse, the transformed alphabet will make

wrududg luug lugu u buzzludg zubblu uv ruddub luddurz. Wu

writing look like a puzzling supply of random letters. We

zuzd uvu udu buzur gwuuzduud: uz ud zuzduvuud? Ugzbugd ludz

just have one major question: is it justified? Expect lots

uv gubbluuddz vrub buublu wu gruw ub wud du uld zuzdub.

of complaints from people who grew up with the old system.

Your turn

https://naclo.org/resources/problems/2020/N2020-L.pdf

The following 14-line poem is one of the four sections of a grammar of Sanskrit, an ancient Indian language, written by the 4th-century BCE Indian grammarian Pāṇini. It is called the *Akṣarasamāmnāya* or *Śivasūtras*, and it functions as an ordering of the sounds of the Sanskrit language[1] – like the English "A, B, C…" with some special properties.

1.	*a*	*i*	*u*			*Ṇ*
2.				*ṛ*	*ḷ*	*K*
3.		*e*	*o*			*Ñ*
4.		*ai*	*au*			*C*
5.	*h*	*y*	*v*	*r*		*Ṭ*
6.					*l*	*Ṇ*
7.	*ñ*	*m*	*ñ*	*ṇ*	*n*	*M*
8.	*jh*	*bh*				*Ñ*
9.			*gh*	*ḍh*	*dh*	*Ṣ*
10.	*j*	*b*	*g*	*ḍ*	*d*	*Ś*
11.	*kh*	*ph*	*ch*	*ṭh*	*th*	
			c	*ṭ*	*t*	*V*
12.	*k*	*p*				*Y*
13.		*ś*	*ṣ*	*s*		*R*
14.	*h*					*L*

NOTE: ṛ and ḷ are vowels; *ñ, m, ṇ, ḍ, ṭ, ś*, and *ṣ* are consonants. A consonant with a letter *h* after it (e.g. *jh*) is considered a separate 'sound' from the consonant without the *h* (e.g. *j*). The vowels *a i u* each have a long counterpart, *ā ī ū*, which for purposes of the *Śivasūtras* is considered equivalent with the short form.

The organization of the *Śivasūtras* allows us to give names to certain groups of sounds. For example, the single syllable *aC* refers to the vowels (*a i u ṛ ḷ e o ai au*). Similarly, *haL* refers to the consonants (all the sounds that are not vowels), and *yaṆ* refers to a specific class of consonants (*y v r l*). Each of these single-syllable words (and the group of sounds that it describes) is known as a *pratyāhāra*.

L1. To what do the following *pratyāhāras* refer? List the sounds:

...iK?

...haṆ?

...khaY?

Solution to part 1

1.	a	i	u			Ṇ
2.				r̥	l̥	K
3.		e	o			N̄
4.		ai	au			C
5.	h	y	v	r		Ṭ
6.					l	Ṇ
7.	ñ	m	ñ	ṇ	n	M
8.	jh	bh				Ñ
9.			gh	ḍh	dh	Ṣ
10.	j	b	g	ḍ	d	Ś
11.	kh	ph	ch	ṭh	th	
			c	ṭ	t	V
12.	k	p				Y
13.		ś	ṣ	s		R
14.	h					L

iK (high vowels and semivowels)

aC (vowels)

haṆ (humming and buzzing sounds)

khaY (voiceless stops)

haL (consonants)

Sound changes

Sounds often change to become more similar to their neighbors in a word.

possible	impossible	**front (m)**
balanced	imbalanced	
temperate	intemperate	**middle (n)**
definite	indefinite	
cautious	incautious	**back (spelled n, pronounced ng)**
gratitude	ingratitude	

Rule: jhaLaḥ jaŚ jhaŚi

jhaL sounds (stops, s sounds and h)

become

jaŚ sounds (voiced stops without h)

when followed by

jhaŚ sounds (voiced stops)

jagatdhana >

tatduḥkha >

bhrātṛnāman >

1.	*a*	*i*	*u*			*Ṇ*	
2.				*ṛ*	*ḷ*	*K*	
3.		*e*	*o*			*Ñ*	
4.		*ai*	*au*			*C*	
5.	*h*	*y*	*v*	*r*		*Ṭ*	
6.					*l*	*Ṇ*	
7.	*ñ*	*m*	*ñ*	*ṇ*	*n*	*M*	
8.	*jh*	*bh*				*Ñ*	jhaŚ
9.			*gh*	*ḍh*	*dh*	*Ṣ*	
10.	*j*	*b*	*g*	*ḍ*	*d*	*Ś*	jaŚ
11.	*kh*	*ph*	*ch*	*ṭh*	*th*		
			c	*ṭ*	*t*	*V*	
12.	*k*	*p*				*Y*	
13.		*ś*	*ṣ*	*s*		*R*	
14.	*h*					*L*	jhaL

Rule: jhaLaḥ jaŚ jhaŚi

jhaL sounds (stops, s sounds and h)
become
jaŚ sounds (voiced stops without h)

when followed by

jhaŚ sounds (voiced stops)

jagatdhana > jaga**d**dhana

tatduḥkha > ta**d**duḥkha

bhrātṛnāman > bhrātṛnāman

1.	*a*	*i*	*u*			*Ṇ*	
2.				*ṛ*	*ḷ*	*K*	
3.		*e*	*o*			*Ñ*	
4.		*ai*	*au*			*C*	
5.	*h*	*y*	*v*	*r*		*Ṭ*	
6.					*l*	*Ṇ*	
7.	*ñ*	*m*	*ṅ*	*ṇ*	*n*	*M*	
8.	*jh*	*bh*				*Ñ*	jhaŚ
9.			*gh*	*ḍh*	*dh*	*Ṣ*	
10.	*j*	*b*	*g*	*ḍ*	*d*	*Ś*	jaŚ
11.	*kh*	*ph*	*ch*	*ṭh*	*th*		
			c	*ṭ*	*t*	*V*	
12.	*k*	*p*				*Y*	
13.		*ś*	*ṣ*	*s*		*R*	
14.	*h*					*L*	jhaL

One more problem

https://naclo.org/resources/problems/2021/N2021-J.pdf

Old Chinese (Zhou)	Middle Chinese (Tang)	English Translation
(1)	*dó*	‘to come to’
mˤə	*mō*	‘soot’
rajs	*ljè*	‘to revile’
pˤək	**(2)**	‘north’
pˤat	**(3)**	‘to stop in the open’
lˤep	*dẽp*	‘butterfly’
(4)	*bãk*	‘calm, still’
dzak	*dzjẽk*	‘stone’
braj	*bjē*	‘to exhaust’

Tips for solving

- Make a chart!
- List all the changes you see…
  - What sounds do they affect?
  - What sounds are next to them?
  - Are there similarities that group these together?
- Test your patterns against the data to find exceptions.

Solution

- **Initial consonants:**
  - r becomes l
  - l becomes d
  - r after a consonant disappears
  - pharyngealization disappears
- **Vowels:**
  - firstly, -aj- loses its final -j (as stated in the introduction)
  - if the initial consonant was pharyngealized, -ə- becomes -o; no change to other vowels
  - otherwise, -a- and -e- become -je-, while -ə- becomes -i
- **Final consonants:**
  - final -t, -k, and -p cause entering tone
  - final -(C)s disappears (along with any consonants before it), leaving departing tone (`)
  - final -' disappears, leaving rising tone (´)
  - otherwise, level tone is used (¯)